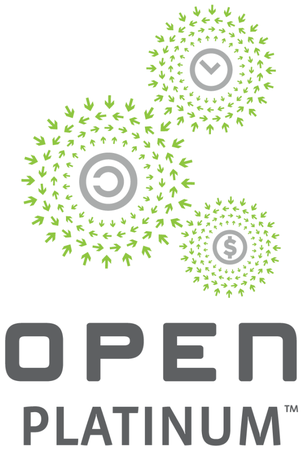

Open. Together.
OCP
SUMMIT

NIC Safe Mode

Yuval Itkin
Distinguished Architect
Mellanox Technologies



Cost optimized servers restrictions

SERVER

In many cases, the only connections to a server are:

1. Power connection
2. Network cable

Mandates server to always be remotely accessible and manageable

Specifications

Cost-optimized server constraints

SERVER

1. Standard platforms are not designed for device-specific customization
2. NIC in OCP cost-optimized server is a single-point-of-failure
   - Bad things happening to a NIC require self-healing method

A need for NIC Safe Mode

When badly-configured NIC prevents a server boot, alternatives are limited

1. Allow modifying bad-configuration via the BMC
   - Not all system settings which could cause such an issue are configurable from the BMC
2. Physically replace of the misconfigured NIC
   - The last resort in a large data center
3. Or….

Examples

```
276-Option Card Configuration Error. An option card is requesting more memory
mapped I/O than is available.
Action: Remove the option card to allow the system to boot.
```

```
 RSP <ffff8808748a3e10>
---[ end trace d553ef6929e1a94a ]---
Kernel panic - not syncing: Fatal exception
Pid: 1, comm: swapper Tainted: G      D    -- ------------    2.6.32-573.el6.x86
_64 #1
Call Trace:
 [<ffffffff81537a84>] ? panic+0xa7/0x16f
 [<ffffffff8153c864>] ? oops_end+0xe4/0x100
 [<ffffffff81010f5b>] ? die+0x5b/0x90
 [<ffffffff8153c094>] ? do_trap+0xc4/0x160
 [<ffffffff8100ef55>] ? do_invalid_op+0x95/0xb0
 [<ffffffff81c6b1fe>] ? pci_assign_unassigned_resources+0xee/0x21d
 [<ffffffff812acd0c>] ? pci_bus_write_config_word+0x6c/0x80
 [<ffffffff812d2f1d>] ? __pci_setup_bridge+0x8d/0x320
 [<ffffffff8100c01b>] ? invalid_op+0x1b/0x20
 [<ffffffff81c6b1fe>] ? pci_assign_unassigned_resources+0xee/0x21d
 [<ffffffff81c78842>] ? pcibios_assign_resources+0x0/0x76
 [<ffffffff81c788b4>] ? pcibios_assign_resources+0x72/0x76
 [<ffffffff810020d0>] ? do_one_initcall+0xc0/0x280
 [<ffffffff81c37a77>] ? kernel_init+0x29b/0x2f7
 [<ffffffff8100969d>] ? __switch_to+0x7d/0x340
 [<ffffffff8100c28a>] ? child_rip+0xa/0x20
 [<ffffffff81c377dc>] ? kernel_init+0x0/0x2f7
 [<ffffffff8100c280>] ? child_rip+0x0/0x20
```

Mellanox NIC Safe Mode benefits

- Needed when a bad configuration of devices prevents a system from starting
- Safe-Mode capability allows device recovery without having to remove it
- Safe-Mode entry is ***Automatic***

Mellanox NIC Safe Mode description

- Supporting devices detect system reset which was not followed by a driver-start
- Upon a pre-configured number of bad reboot cycles, ***ConnectX*** device automatically enters Safe Mode
- Safe Mode can be enabled/disabled through non-volatile configuration
- Safe Mode can be enabled/disabled/monitored through HII, NC-SI and the OS using Mellanox tools

Mellanox NIC Safe Mode operation

- Device in Safe Mode requires minimal system resources to allow the system to start
- Once operating in Safe Mode, bad settings can be reviewed & modified by the user/operator
- Device operating in Safe Mode provides visibility to its operating mode through HII, Console and through Mellanox configuration tools

```
[root@qahp-104 ~]# dmesg | grep safe
mlx5_core 0000:04:00.0: 0000:04:00.0:mlx5_cmd_init_hca:230:(pid 2783): Warning: Device is operating in safe mode after 3 bad boots, settings minimized
 to allow operational configuration
mlx5_core 0000:04:00.1: 0000:04:00.1:mlx5_cmd_init_hca:230:(pid 2837): Warning: Device is operating in safe mode after 3 bad boots, settings minimized
 to allow operational configuration
```

- After reconfiguring the device to the correct settings, it will restart normally with the new settings on the next system reboot

Mellanox NIC Safe Mode configuration options

4 different operational modes are possible with Mellanox NIC Safe Mode

1. NIC Safe Mode is disabled
2. NIC Safe Mode is enabled after *Num-Bad-Reboots* (default mode)
3. NIC Safe Mode is activated once in the next reboot
4. NIC Safe Mode is enforced for any boot
   - Safe Mode default can be set to disabled/enabled, through non-volatile configuration
   - *Num-Bad-Reboots* parameter can be between 1-255 bad reboots

Call for action

- Request new NC-SI standard command to force NIC to reset to “factory default” mode
- When in “factory default” mode, a given device shall always allow re-configuration

Specifications


Open. Together.
OCP Global Summit | March 14–15, 2019
OCP
SUMMIT